

Warm season heavy rainfall events over the Huaihe River Valley and their linkage with wintertime thermal condition of the tropical oceans

Laifang Li^{1,5} · Wenhong Li¹ · QiuHong Tang² · Pengfei Zhang^{3,4} · Yimin Liu³

Received: 13 September 2014 / Accepted: 16 March 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Warm season heavy rainfall events over the Huaihe River Valley (HRV) of China are amongst the top causes of agriculture and economic loss in this region. Thus, there is a pressing need for accurate seasonal prediction of HRV heavy rainfall events. This study improves the seasonal prediction of HRV heavy rainfall by implementing a novel rainfall framework, which overcomes the limitation of traditional probability models and advances the statistical inference on HRV heavy rainfall events. The framework is built on a three-cluster *Normal* mixture model, whose distribution parameters are sampled using Bayesian inference and Markov Chain Monte Carlo algorithm. The three rainfall clusters reflect probability behaviors of light, moderate, and heavy rainfall, respectively. Our analysis indicates that heavy rainfall events make the largest contribution to the total amount of seasonal precipitation.

Furthermore, the interannual variation of summer precipitation is attributable to the variation of heavy rainfall frequency over the HRV. The heavy rainfall frequency, in turn, is influenced by sea surface temperature anomalies (SSTAs) over the north Indian Ocean, equatorial western Pacific, and the tropical Atlantic. The tropical SSTAs modulate the HRV heavy rainfall events by influencing atmospheric circulation favorable for the onset and maintenance of heavy rainfall events. Occurring 5 months prior to the summer season, these tropical SSTAs provide potential sources of prediction skill for heavy rainfall events over the HRV. Using these preceding SSTA signals, we show that the support vector machine algorithm can predict HRV heavy rainfall satisfactorily. The improved prediction skill has important implication for the nation's disaster early warning system.

✉ Wenhong Li
wenhong.li@duke.edu

¹ Earth and Ocean Sciences, Nicholas School of the Environment and Earth Sciences, Duke University, 321C Old Chem. Bldg, P.O. Box 90227, Durham, NC 27708, USA

² Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, People's Republic of China

³ State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, People's Republic of China

⁴ University of Chinese Academy of Sciences, Beijing, People's Republic of China

⁵ Present Address: Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

Keywords Heavy rainfall events · Seasonal climate prediction · Bayesian inference on precipitation · Huaihe River Valley · *Normal* mixture model · Support vector machine

1 Introduction

The Huaihe River Valley (HRV) is a key agricultural area in China to sustain food supplies to whole nation, where summertime flooding events are a primary breaker of local agriculture (Li et al. 2011). Historical records showed that devastating flooding events over this region are tightly associated with heavy rainfall events. For example, in the late June and early July of 2003, consecutive heavy rainfall events brought about 450 mm precipitation into the HRV, causing severe flooding, loss of life and billions of dollars in agricultural and economic damage (Zhang and

You 2014). Thus, accurately predicting HRV heavy rainfall events is of vital importance to reduce the economic loss caused by flooding and downpours.

However, previous studies on rainfall prediction focused primarily on the seasonal mean precipitation, which might not be readily adaptable to heavy rainfall events. Specifically, the statistical behavior of heavy rainfall events are determined by the tail of probability distributions (Katz and Brown 1992), which is substantially different from that of the mean (Gumbel 1954; Wilson and Toumi 2005). Furthermore, it has been noticed that heavy rainfall events respond differently to climate variability and climate change compared to the seasonal mean precipitation (Chen et al. 2012; Chou et al. 2012; Qian et al. 2007; Wu and Fu 2013). Consequently, the factors responsible for seasonal mean precipitation over the HRV, including the sea surface temperature anomaly (SSTA) patterns (Chang et al. 2000; Huang and Sun 1992; Yang and Lau 2004), springtime soil moisture content (Meng et al. 2014; Zhang and Zuo 2011), and Arctic sea ice extent (Li and Leung 2013; Vihma 2014), might not be used in the same way as for the prediction of heavy rainfall events.

In order to improve seasonal prediction of heavy rainfall events, the prediction models should be updated and the climate predictors suitable for the HRV heavy rainfall events should be identified. On top of that, a reliable statistical inference on the HRV heavy rainfall events and a better understanding of the related physical processes are needed.

This study advances climate prediction of HRV heavy rainfall by applying a novel statistical framework. Unlike traditional statistical prediction models, the framework does not require predefinition of distribution kernels, and thus largely reduces the biases in rainfall statistical models due to their subjective selection of distribution kernels (Li and Li 2013). By objectively identifying heavy rainfall events, the framework can better model the statistical behavior of heavy rainfall, and thus help to improve the understanding of related physical mechanisms. The probability behavior of heavy rainfall events are then linked to preceding SSTA pattern to identify the potential climate predictors. Two statistical prediction models, one linear model and the other machine learning algorithm, are constructed to assess the predictability of the HRV heavy rainfall events using the identified SSTA predictors.

The rest of the paper is organized as following. In Sect. 2, data, analysis methods, and the statistical framework are described. The Bayesian inference on the HRV summertime heavy rainfall events is presented in Sect. 3. The SSTA predictors of the HRV heavy rainfall are identified based on exploratory data analysis in Sect. 4. Furthermore, the atmospheric circulation is analyzed to create the physical linkage between preceding SSTA and summertime

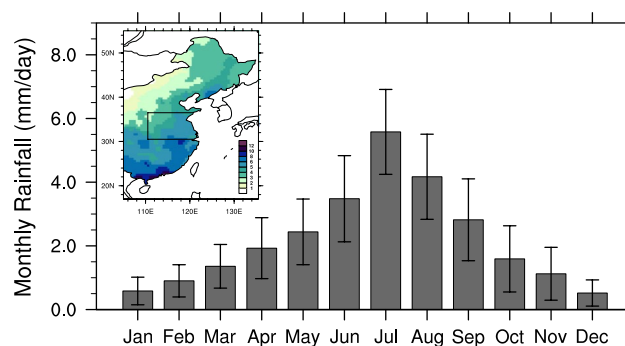


Fig. 1 Climatology (1961–2012) of monthly mean precipitation (gray bars, mm day^{-1}) over the HRV; the error bars denote one standard deviation of interannual variation of precipitation in each month. The inner plot shows the 1961–2012 climatology of summer (JJA) precipitation in Eastern China. The black box denotes the HRV region

HRV heavy rainfall events. In Sect. 5, two statistical prediction models, multiple linear regression model and support vector machine (SVM) algorithm, are constructed to test the predictability of the HRV heavy rainfall events using the SSTA predictors identified in Sect. 4. Concluding remarks are presented in the last section.

2 Data and methods

2.1 Data

The data used in this study includes gridded daily precipitation data archived by China Meteorological Administration (CMA) (Xie et al. 2007). The temporal period analyzed in this study is 1961–2012 in order to avoid the potential inaccurate of statistical inference due to the sparse data coverage prior to 1960s. The HRV covers a geographical domain of 30.5°N – 36.5°N ; 110.5°E – 121.5°E (Hong and Liu 2012), which is delineated by the box in Fig. 1. Daily precipitation within this domain is averaged in order to capture the regional-scale features of rainfall from the climate perspective.

Summer season is defined as June–July–August (JJA), when precipitation over the HRV peaks (Fig. 1). The precipitation during these three months also displays the strongest interannual variation, indicating the importance of JJA precipitation to climate variability in this region (Fig. 1). Furthermore, summer precipitation in the HRV is critical to local agriculture and economy, making warm-season rainfall a central problem of climate prediction for this region.

Atmospheric circulation fields, i.e. wind patterns, were adopted from the National Center of Environment Prediction/National Center of Atmospheric Research (NCEP/

NCAR) reanalysis (Kalnay et al. 1996) during 1961–2012. In this study, both daily and JJA mean circulation were analyzed. In order to explore the potential predictability of rainfall, both synchronized and preceding sea surface temperature anomalies (SSTA) were examined. The SST data is from the National Oceanic and Atmospheric Administration (NOAA) Extended Reconstructed SST [ERSST Version 3 (Smith et al. 2008)].

2.2 Rainfall probability framework: finite Normal mixture model

To describe the probability distribution of summer rainfall over the HRV, a rainfall framework based on a three-cluster *Normal* mixture model was implemented (Li and Li 2013). The advantages of the *Normal* mixture model lie in its flexibility of distribution shapes, because any smoothed distribution can be approximated by the combination of a finite number of *Normals*. Thus, it overcomes the limitation of traditional statistical models that requires the subjective selection of predefined distribution kernels and cannot be easily adapted to different climate zones. For example, the *Log-Normal* distribution tends to better model the rainfall in the subtropical regions, while the *Gamma* distribution fits tropical precipitation better (Cho et al. 2004). Such a subjective selection of distribution models can introduce bias to the statistical inference on the HRV heavy rainfall events. The flexibility of *Normal* mixture model in distribution shape is especially important for HRV summer rainfall, because daily rainfall displays multi-modal feature that cannot be captured by traditional models with uni-modal distribution kernels.

When constructing finite *Normal* mixture models, choosing the optimal number of clusters is difficult and sometimes controversial (McLachlan and Peel 2000; Melnykov and Maitra 2010; Richardson and Green 1997). Generally, adding more clusters to the mixture model could better approximate the true distribution of rainfall. However, an unlimited increase in clusters also increases the risk of over-fitting (Lin et al. 2007), blurs the physical meanings of each cluster, and hampers the interpretation of mechanisms that control the rainfall probability distribution. In this study, a three-cluster *Normal* mixture was constructed. The three clusters reflect the probability behavior of different summer rainfall categories over the HRV, i.e. the light, moderate, and heavy rainfall according to the American Meteorological Society (AMS) Glossary of Meteorology (2009).

The three-cluster finite *Normal* mixture model takes the mathematical form:

$$y_i \mid \pi, \mu, \phi \sim \sum_{h=1}^3 \pi_h N(y_i \mid \mu_h, \phi_h^{-1}), \quad (1)$$

where π is the weight of rainfall clusters (i.e. the frequency of each rainfall types). It is noteworthy that $\sum_{h=1}^3 \pi_h = 1$, and thus π_h is mutually dependent. μ is the cluster mean, and ϕ is the precision of *Normal* distributions. $h \in \{1, 2, 3\}$ is the cluster index.

To obtain the distribution parameters (π , μ , and ϕ) in the mixture model, Bayesian statistical inference was implemented. The priors about π , μ , and ϕ are as follows:

$$\pi \mid a_1, a_2, a_3 \sim \text{Dirichlet}(a_1, a_2, a_3) \quad (2)$$

$$(\mu_h, \phi_h) \sim \text{Normal}(\mu_h \mid \mu_{0h}, \kappa_h \phi_h^{-1}) \text{Gamma}(\phi_h \mid \alpha_h, \beta_h) \quad (3)$$

In Eq. (3), κ is the degree of freedom, and the *Gamma*($\phi_h \mid \alpha_h, \beta_h$) was parameterized to have mean α_h/β_h and variance α_h/β_h^2 . The parameters in the prior distributions (Eqs. 2–3) were assigned according to the AMS definitions of light (0–6 mm day^{−1}), moderate (6–18 mm day^{−1}), and heavy rainfall (>18 mm day^{−1}) (AMS Glossary of Meteorology 2009). However, we kept the prior distributions weakly informative to incorporate more data information into the posterior distribution: $(a_1, a_2, a_3) = (0.5, 0.35, 0.15)$, $\mu_{0h} = (1.0, 8.0, 20.0)$, $\kappa_h = (1, 1, 1)$, $\alpha_h = (1.0, 1.0, 0.4)$, and $\beta_h = (1.0, 1.0, 1.0)$.

Since the priors (Eqs. 2–3) and the likelihood model (Eq. 1) are semi-conjugate, the full conditional posterior distributions can be derived analytically (Gelfand 2000). The Gibbs sampler for posterior computation using Markov Chain Monte Carlo (MCMC) algorithm is as follows:

$$\Pr(z_i = h \mid -) = \frac{\pi_h \text{Normal}(y_i \mid \mu_h, \phi_h^{-1})}{\sum_{h=1}^3 \pi_h \text{Normal}(y_i \mid \mu_h, \phi_h^{-1})} \quad (4)$$

$$(\mu_h, \phi_h \mid -) \sim \text{Normal}(\mu_h \mid \hat{\mu}_{0h}, \hat{\kappa}_h \phi_h^{-1}) \text{Gamma}(\phi_h \mid \hat{\alpha}_h, \hat{\beta}_h), \quad (5)$$

where $\hat{\kappa}_h = (\kappa_h^{-1} + n_h)^{-1}$; $\hat{\mu}_{0h} = \hat{\kappa}_h (\kappa_h^{-1} \mu_{0h} + n_h \bar{y}_h)$; $\hat{\alpha}_h = \alpha_h + \frac{n_h}{2}$; and $\hat{\beta}_h = \beta_h + \frac{1}{2} \left\{ \sum_{i=1}^{n_h} (y_i - \bar{y}_h)^2 + \left(\frac{n_h}{1 + \kappa_h n_h} \right) (\bar{y}_h - \mu_{0h})^2 \right\}$. $n_h = \sum_{i=1}^n 1(z_i = h)$ denotes the number of samples in cluster h , and $\bar{y}_h = n_h^{-1} \sum_{i:z_i=h} y_i$ is the sample mean of cluster h .

$$(\pi_1, \pi_2, \pi_3 \mid -) \sim \text{Dirichlet}(a_1 + n_1, a_2 + n_2, a_3 + n_3) \quad (6)$$

In this study, the MCMC algorithm (Eqs. 4–6) is applied to daily rainfall in each summer during the 1961–2012. Since heavy rainfall intensity is stronger than moderate rainfall and light rainfall, this physical constraint is placed upon μ_h ($\mu_1 < \mu_2 < \mu_3$) to deal with the label switching issues (Stephens 2000). The MCMC algorithm is run 1000 times and the first 200 burn-in samples were discarded, and

Table 1 Comparison of distribution models used to describe the probability behaviors of summer precipitation over the HRV

Distribution model	Kernel function	Parameters	Average square error
Gamma	$f(x) \propto x^{\alpha-1} e^{-\beta x}; x \in (0, +\infty)$	$\alpha > 0$: shape parameter $\beta > 0$: rate parameter	35.85
Log-Normal	$f(x) \propto x^{-1} e^{-\frac{1}{2} \left(\frac{\ln(x)-\mu}{\sigma} \right)^2}; x \in (0, +\infty)$	$\mu \in R$: shape parameter $\sigma > 0$: scale parameter	55.43
Weibull	$f(x) \propto x^{k-1} e^{-(\frac{x}{\lambda})^k}; x \in [0, +\infty)$	$k > 0$: shape parameter $\lambda > 0$: scale parameter	37.53
Exponential	$f(x) \propto \lambda e^{-\lambda x}; x \in [0, +\infty)$	$\lambda > 0$: rate parameter	66.99
Normal mixture (Sect. 2.2)	$f(x) \propto \sum_{h=1}^3 \pi_h \phi_h e^{-\frac{1}{2} \phi_h^2 (x-\mu_h)^2}$	$\pi_h > 0; \sum_{h=1}^3 \pi_h = 1$: cluster weight μ_h : cluster mean $\phi_h > 0$: cluster precision	30.64

the remaining 800 post burn-in samples were used in the analysis.

The goodness-of-fit of the constructed three-cluster *Normal* mixture model was compared with the widely used distribution models, including the *Gamma*, *Exponential*, *Weibull*, and *Log-Normal* (Table 1). The quantile-based square error of the five distribution models was calculated to assess the “goodness-of-fit” of each model. The smaller square error indicates a better approximation of models to observed rainfall distribution. According to Table 1, the *Normal* mixture model shows the smallest square error, suggesting its better performance than the other four traditional rainfall models. Thus, the application of the *Normal* mixture model can improve the statistical inference and diagnostic analysis of regional hydroclimate over the HRV.

3 Bayesian inference on summer precipitation over the HRV

From the posterior distributions (Eqs. 4–6), the parameters in the *Normal* mixture model (Eq. 1) can be sampled. In this study, the interannual variation of the distribution parameters μ_h and π_h are of primary interest, where μ_h reflects the intensity of light, moderate, and heavy rainfall and π_h describes their frequency.

3.1 Bayesian inference

Figure 2 shows the year-to-year fluctuations of μ_h and π_h over the HRV region during the period of 1961–2012. The 52-year climatology is about 1.09 mm day^{−1} for light rainfall, 5.45 mm day^{−1} for moderate rainfall, and 16.23 mm day^{−1} for heavy rainfall, respectively (Fig. 2a). Compared to the AMS criteria of the three rainfall types, the μ_h derived from the posterior distribution is lower for moderate and heavy rainfall, indicating that the criteria used to define rainfall categories should be adjusted at

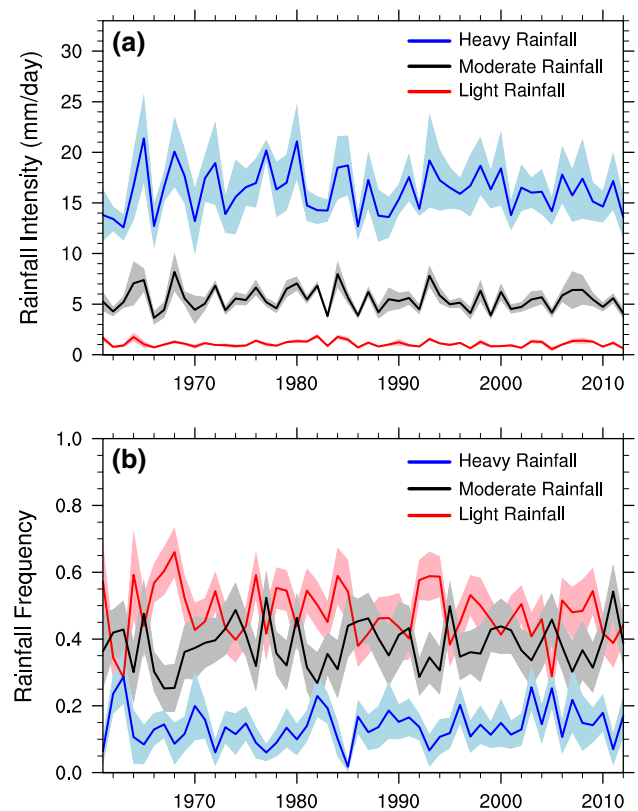
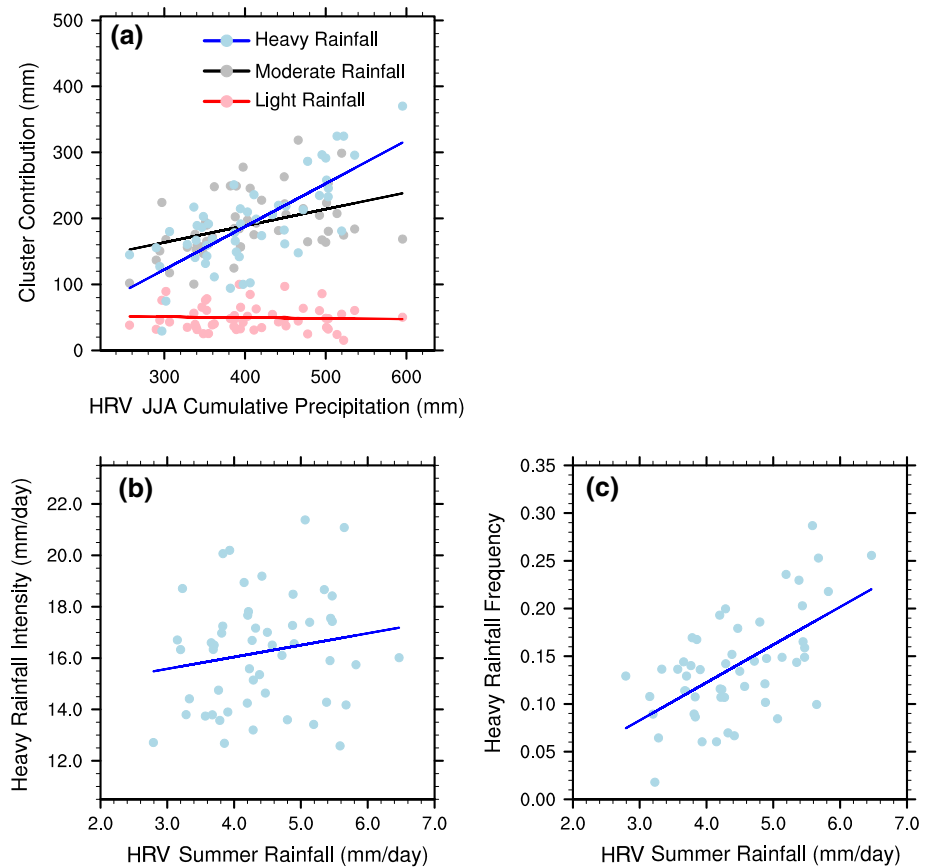


Fig. 2 Bayesian inference on the interannual variation of the **a** intensity (mm day^{−1}) and **b** frequency of light (red curves); moderate (black curves); and heavy (blue curves) rainfall events. Shading represents the 95 % credible interval as derived from the post burn-in Markov Chain Monte-Carlo samples

different regions. Constrains placed on μ_h are needed in the algorithm so that there is no overlap of the sampled μ_h between rainfall clusters in the MCMC algorithm (Fig. 2a). Besides rainfall intensity, the frequency of each rainfall category (π_h) can also be derived from the framework. The climatology of light rainfall frequency is 49 %, moderate rainfall frequency is 38 %, and heavy rainfall frequency is 13 %, respectively (Fig. 2b).

Fig. 3 **a** Contribution of light (red dots), moderate (gray dots), and heavy (blue dots) rainfall clusters to summer season cumulative precipitation over the HRV (mm); and the relationship between summer precipitation (unit: mm day⁻¹) and **b** heavy rainfall intensity (blue dots; unit: mm day⁻¹) and **c** heavy rainfall frequency (blue dots). The straight lines are the best least square fitting lines



The sampling uncertainty of μ_h and π_h from the MCMC algorithm is also calculated. The uncertainty range is derived as the 95 % credible interval¹ based on post burn-in MCMC samples (Hoff 2009). Comparatively, the inference shows the highest confidence of light rainfall intensity among all three clusters, due to relatively larger sample size (\hat{k}) and smaller sample standard deviation (ϕ_1^{-1}) in this cluster. In contrast, the heavy rainfall cluster has the least rainfall samples, and the cluster mean μ_3 usually shows largest uncertainty compared to μ_1 and μ_2 (Fig. 2a). However, compared to other regions such as the Southeastern United States (Li and Li 2013), sampling uncertainty over the HRV is within the magnitude of interannual variation of heavy rainfall intensity in the period of 1961–2012, i.e. the inference on the HRV heavy rainfall events is of considerably higher confidence.

It is noteworthy that Bayesian inference on *Normal* mixture model does not require the sample sets to be strictly independent and identical distributed (i.i.d.) (Hoff 2009). In other words, the inference on posterior distribution is not impacted by the autocorrelation of rainfall samples.

We run the model using de-autocorrelated samples and the inference on the distribution parameters is almost identical (Appendix 1). The results suggest that the model can be applied to climate variables, including precipitation and temperature, which usually have high autocorrelation. In the following study, the inference using the original sample set is adopted for better temporal consistency.

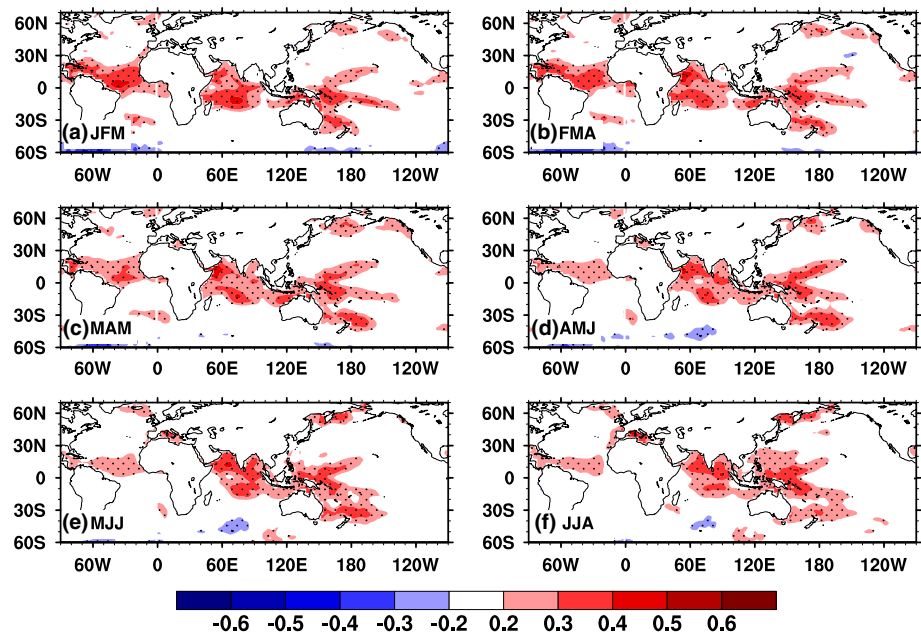
3.2 Contribution of heavy rainfall to HRV regional hydroclimate

According to the *Normal* mixture model, the sample mean (i.e. seasonal mean) of rainfall equals to the weighted average of the three rainfall clusters: $\bar{Y}|\pi, \mu, \phi = \sum_{h=1}^3 \pi_h \mu_h$. Utilizing such a relationship, the contribution of each rainfall cluster to summertime hydroclimate over the HRV can be assessed.

Figure 3a shows the relationship between $\pi_h \mu_h$ and seasonal mean precipitation (\bar{Y}). Climatologically, the HRV receives 432 mm precipitation in JJA, with approximately 50 mm from light rainfall, 190 mm from moderate rainfall, and 192 mm from heavy rainfall (Fig. 3a). The contributions of moderate and heavy rainfalls to total seasonal precipitation amount are almost equivalent. However, the variance of seasonal mean precipitation is primarily explained

¹ Credible interval in Bayesian statistics is equivalent to confidence interval in Frequentist statistics.

Fig. 4 Sea surface temperature anomalies (SSTA; shaded) regresses upon the interannual variation of heavy rainfall frequency over the HRV: **a** JFM, **b** FMA, **c** MAM, **d** AMJ, **e** MJJ, and **f** JJA. The regression coefficients significant at $\alpha = 0.05$ level are stippled



by the heavy rainfall at interannual scales. Specifically, the R^2 between heavy rainfall and seasonal mean precipitation reaches 0.58 for the 1961–2012 period, indicating that about 60 % of the JJA mean precipitation variance can be explained by the heavy rainfall events (Fig. 3a). In contrast, the light and moderate rainfalls together explain only 21 % of the seasonal rainfall variance (Fig. 3a). The results indicate that heavy rainfall events can significantly modulate seasonal scale hydroclimate over the HRV, exerting persistent climatic impact on regional water resources.

The contributions of heavy rainfall events to summer season precipitation are mainly through the interannual variation of rainfall frequency (π_3) rather than heavy rainfall intensity (μ_3). Specifically, the correlation between heavy rainfall intensity (μ_3) and seasonal mean precipitation (\bar{Y}) is 0.17, statistically insignificant (Fig. 3b). In contrast, the frequency of heavy rainfall is highly correlated with seasonal mean precipitation, with correlation coefficient approaching 0.60 ($p < 0.001$). Thus, in order to achieve a better seasonal prediction of HRV summer precipitation, improvements in the understanding of factors and processes responsible for HRV heavy rainfall frequency is a key.

4 Wintertime SSTA related to HRV heavy rainfall events in the summer

Previous studies have suggested that global climate modes could influence the rainfall events by changing the large-scale background circulation where synoptic-scale systems develop (Higgins et al. 2007; Ropelewski and Halpert 1987; Ting and Wang 1997). In the East China, summer rainfall

events are associated with the East Asia summer monsoon (EASM) (Ding 1992; Lau 1992; Matsumoto 1988), and are further linked to the well-defined climate modes, including the El Niño–Southern Oscillation (ENSO) (Wang et al. 2003; Wu and Wang 2002), the Pacific Decadal Oscillation (PDO) (Lei et al. 2011), and the Indian Ocean Dipole (IOD) (Yang et al. 2010), as well as anthropogenic forcing (Wang et al. 2013).

The linkage between the HRV heavy rainfall frequency and climate factors is explored by regressing SSTA upon the sampled heavy rainfall frequency. Establishing such linkages could provide insights for seasonal prediction of HRV heavy rainfall events as well as summertime hydroclimate over the HRV.

4.1 Association of tropical SSTA to heavy rainfall events over the HRV

Figure 4 shows the precedent and synchronized SSTA regressed upon the frequency of the heavy rainfall events. Generally, the increased frequency of HRV heavy rainfall events is associated with warmer SSTA over the north Indian Ocean, the equatorial western Pacific, and the tropical Atlantic Ocean (Fig. 4). Specifically, the warm SSTAs over the three tropical oceans occur in the preceding winter (January–February–March, JFM), 5 months before the summer (JJA, Fig. 4a). In the north Indian Ocean and the equatorial western Pacific, SSTAs persist throughout summer (Fig. 4b–f), indicating that the SSTAs might influence HRV heavy rainfall events by providing persistent oceanic forcing on the overlying atmosphere and thus impacting the summertime circulation pattern over the HRV.

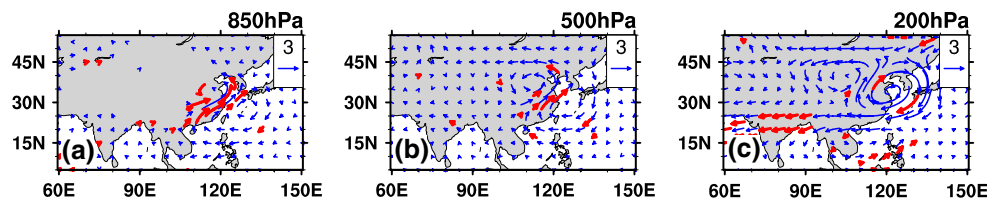


Fig. 5 Typical atmospheric circulation (vectors, m s^{-1}) patterns (represented as wind anomalies) composite upon heavy rainfall events as identified using posterior predictives (Eqs. 4–6): **a** 850 hPa,

b 500 hPa, and **c** 200 hPa. The bold red vectors represent wind anomaly significant at 95 % confidence interval

Compared to the SSTAs over the Indian Ocean and the equatorial western Pacific, the tropical Atlantic SSTA is less persistent. In JFM, the warm SSTAs span the entire tropical Atlantic and the Intra-Americas Seas (Fig. 4a). In the following months, the areas with significant SSTAs gradually shrink and the SSTAs in the southern portion of the tropical oceans completely decay in JJA months (Fig. 4d–f). In the north Tropical Atlantic, although statistically significant, the SSTAs become weaker in summer months (Fig. 4f). The gradually decaying feature of SSTAs in the tropical Atlantic may be related to the negative feedback between tropical oceans and the seasonal evolution of atmospheric circulation, including North Atlantic Oscillation (Huang and Shukla 2005). Thus, tropical Atlantic SSTA might influence summertime atmospheric circulation through the air–sea interaction over the Atlantic and modulate heavy rainfall events over the HRV (Sun and Wang 2012).

In conclusion, the analysis of HRV heavy rainfall frequency and global SSTA identified a significant relationship between summertime HRV heavy rainfall events and wintertime tropical SSTAs. It is noteworthy that the above-mentioned relationship should not be interpreted as a coincidence between the SST warming trend and the wetting trend over the HRV. Regression analysis using detrended SSTA records shows similar results, except minor changes in the regression coefficients (not shown). These tropical SSTAs occur 5 months prior to summer season, providing potential sources of predictability for the HRV heavy rainfall frequency.

4.2 Contribution of tropical SSTA to the typical circulation pattern associated with HRV heavy rainfall

The analysis of HRV heavy rainfall and global SSTA identifies a positive relationship between tropical SSTAs and the HRV heavy rainfall frequency. Usually, the SSTA influences heavy rainfall events by changing the large-scale atmospheric circulation that provides background conditions for the development of rain-bearing systems. Thus, exploring the contribution of tropical SSTAs to the

large-scale circulation patterns favorable for the HRV heavy rainfall events can improve the mechanistic understanding of the relationship between tropical SSTAs and the HRV heavy rainfall events.

The typical circulation pattern associated with the HRV heavy rainfall is analyzed based on the posterior predictive. Specifically, using the post burn-in parameter sets (i.e. the 800 MCMC samples), the probability of individual rainfall events falling into different rainfall clusters can be quantified (Eq. 4). By comparing the calculated probability, we can objectively categorize each rainfall event and identify all heavy rainfall events during 1961–2012. Thus, the atmospheric circulation patterns associated with the heavy rainfall events can be obtained accordingly.

Figure 5 shows the circulation maps at 850, 500, and 200 hPa composite upon the objectively identified heavy rainfall events. Climatology of the circulation has been removed from each composite map. According to Fig. 5, heavy rainfall events are associated with abnormally strong southerly wind over Eastern China at the 850 hPa level (Fig. 5a). The significantly intensified southerly wind indicates a strong East Asia summer monsoon, which favors the onset of heavy rainfall events by increasing the moisture supplies from the South China Sea (SCS) (Zhou and Yu 2005) and along the western edge of western Pacific subtropical high (Zhang 2001).

At 500 hPa, an anomalous anticyclone occurs with its center located off the eastern coast (Fig. 5b). The anticyclone is associated with an intensification and southwestward movement of the northwestern Pacific subtropical high. Accompanying this anticyclone, a cyclone is generated inland, presenting a meridionally oriented circulation pattern (Fig. 5b). Such a circulation pattern is consistent with previous studies (Bao 2008).

This meridionally oriented circulation is also observed in the upper troposphere (200 hPa), although the upper tropospheric circulation pattern becomes less significant and shifts northward by about 5 degrees (Fig. 5c). The northward tilt of the circulation pattern in the vertical direction indicates an anomalously strong EASM, favoring the northward migration of monsoon rain-belt to the HRV region.

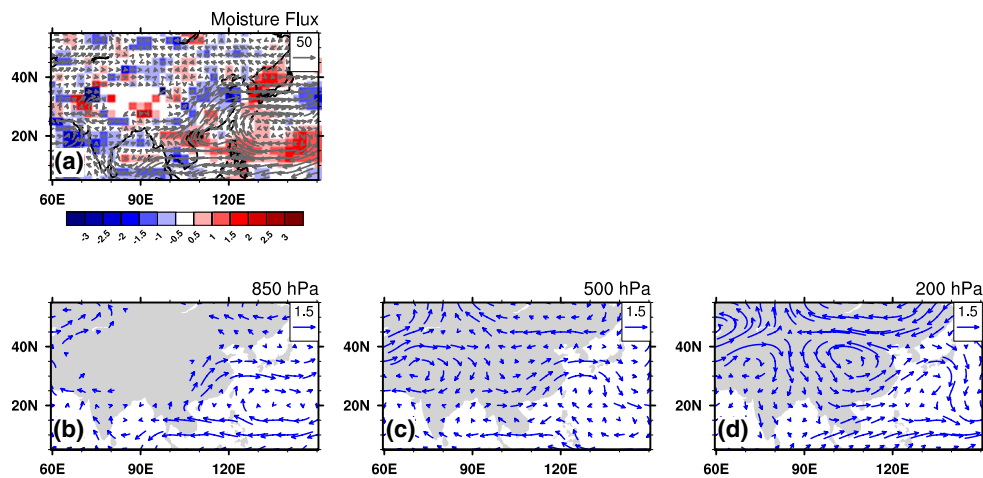


Fig. 6 Differences field of JJA moisture flux (vector, unit: $\text{kg m}^{-1} \text{s}^{-1}$) and moisture flux divergence (shaded, unit: mm day^{-1}) between warm and cold tropical SSTAs (**a**); **b–d** different fields in 850 hPa, 500 hPa, and 200 hPa wind (blue vectors; unit: m s^{-1}),

respectively. The years with JFM SSTA over all three regions above 0.5 (below -0.5) standard deviation are selected to denote the warm (cold) events

These typical circulation patterns tend to become more frequent when the tropical oceans are warmer, according to the composite analysis of JJA circulation upon the JFM SSTAs over the north Indian Ocean, the equatorial western Pacific, and the tropical Atlantic² (Fig. 6). The warm years are defined as when the SSTAs exceed 0.5 standard deviation (STD); while cold years are when SSTAs are below -0.5 STD. Here, 0.5 STD instead of 1 STD criterion is used in order to increase the size of the composite samples.

According to the composite analysis, SSTAs over the tropical oceans collectively contribute to the atmospheric circulation patterns favorable for the HRV heavy rainfall events (Figs. 5, 6). Specifically, warmer tropical oceans intensify northward moisture transport into the Eastern China, which enhances moisture supplies from the SCS and along the western ridge of the western Pacific subtropical high (Zhang 2001; Zhou and Yu 2005). The anomalous moisture flux converges in the HRV, facilitating the heavy rainfall events there (Fig. 6a). The intensification of northward moisture flux is associated with an increase in 850 hPa southerly wind (Fig. 6b), a key circulation feature to sustain HRV heavy rainfall events (Fig. 5a). Furthermore, the southerly wind can enhance the upward motion as indicated by Sverdrup vorticity balance ($\beta v \propto -f \frac{\partial \omega}{\partial z}$) along the subtropics (Liu et al. 2004, 2007).

Intensification of the southerly wind along with the anomalous easterly wind over the SCS (Fig. 6b) might result from the Gill-type response of atmospheric circulation to the warming over the north Indian Ocean (Gill 1980). The warming of the north Indian Ocean generates equatorial Kelvin waves, which enhance tropical easterlies and induces southerly wind to the east of the SSTA center (Gill 1980; Kosaka et al. 2013). Such a circulation pattern as shown in Fig. 6b is consistent with the numerical simulation of tropical Indian Ocean's impact on the HRV summer precipitation (Hong and Liu 2012). Furthermore, the warming over the equatorial western Pacific can reinforce the southerly wind through a meridionally oriented tropical–extratropical teleconnection pattern. The warming over the equatorial western Pacific induces a cyclone over the local SSTA center and an anticyclone off the eastern coast of China (Huang and Sun 1992; Ji et al. 2014; Kosaka and Nakamura 2010). The anticyclone can enhance the southerly wind over the East China, favoring the HRV heavy rainfall events. In addition, the tropical Atlantic also contribute to the above mentioned atmospheric circulation, likely through the wave propagation (Zuo et al. 2013).

In the 500 hPa, the warming over the tropical oceans is accompanied by the intensification of the Northwest Pacific subtropical high and the westward extension of its western ridge (Fig. 6c), consistent with previous studies (Li et al. 2012; Zhou et al. 2009). The extension of the western ridge is a typical feature of the intensified EASM (Chang et al. 2000), which provides a favorable circulation pattern for the HRV heavy rainfall events (Fig. 5b). In the upper troposphere (200 hPa), the circulation composite on tropical SSTAs also resembles the typical patterns during the HRV heavy rainfall events (Fig. 5c, d).

² The North Indian Ocean SSTA index is calculated as the area-averaged SSTA over the 20°S – 25°N , 30°E – 120°E . The equatorial western Pacific SSTA is defined as the averaged SSTA over the 30°S – 10°N , 140°E – 175°E , where only the grid points with climatological SST higher than 300 K were included in the calculation. Tropical Atlantic SSTA is averaged over the 10°S – 25°N , 80°W – 5°W .

Overall, the above analysis suggests that a collective impact of tropical SSTAs on HRV heavy rainfall events, which is likely through the oceanic influence on large-scale circulation. Warmer tropical oceans can enhance the southerly wind and supply excessive moisture to the HRV, favorable for the generation and maintenance of heavy rainfall events.

5 Predictability of HRV heavy rainfall frequency

The analysis of HRV heavy rainfall, atmospheric circulation, and SSTA patterns demonstrates a significant relationship between heavy rainfall frequency and tropical SSTA in proceeding months (Fig. 4a–e). The tropical SSTAs occur five months before the HRV rainy season (Fig. 4f), providing potential sources of predictability for heavy rainfall events over this region.

Using these tropical SSTAs as predictors, statistical prediction models are constructed to predict HRV heavy rainfall events. In this study, multiple linear regression model and support vector machine (SVM) algorithm are applied to assess whether the selection of prediction models contribute to the prediction skill of HRV heavy rainfall.

The multiple linear regression aims to model the relationship between two or more predictors and one response variable by fitting a linear equation to observed data (Christensen 2011). In this study, the regression model is formulated as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon, \quad (7)$$

where y is the observed heavy rainfall frequency over the HRV; x_1 , x_2 and x_3 denote the SSTAs over the north Indian Ocean, equatorial western Pacific, and the tropical Atlantic, respectively; and ε is the residual of the prediction model. β_0 , β_1 , β_2 and β_3 are derived so that the least-square errors between observation and prediction model is minimized.

Consistent with the exploratory data analysis, multiple linear regression using tropical SSTAs shows certain skill in predicting heavy rainfall frequency over the HRV (Fig. 7). The R^2 between prediction and observation is 0.20, significant at $\alpha = 0.01$ level. However, the model (Eq. 7) underestimates the heavy rainfall frequency at the distribution tails (Fig. 7). Thus, it is likely that nonlinearity between predictors and response variable should be considered in order to achieve a better prediction skill (Fig. 7).

The potential nonlinearity between predictors and response variable is considered by constructing SVM regression model (Burgess 1998). The SVM algorithm uses nonlinear kernel function and is formulated as:

$$\hat{y} = f(x_1, x_2, x_3) = w^T \Phi(x_1, x_2, x_3) + b, \quad (8)$$

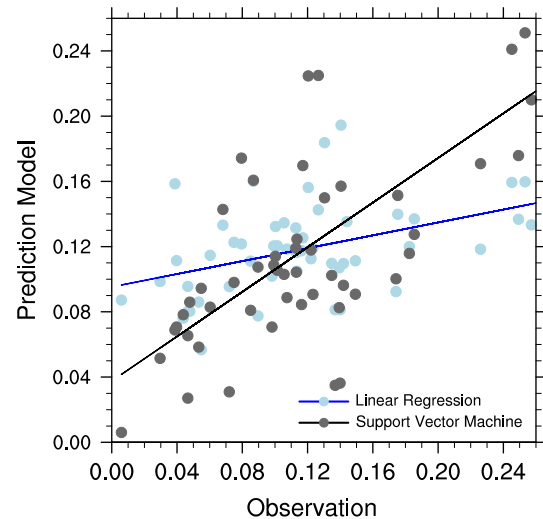


Fig. 7 Heavy rainfall frequency over the HRV as predicted by multiple linear regression model (blue dots) and support vector machine (SVM) algorithm (gray dots). The straight lines are the best least squares fitting lines between observations and prediction models

where (x_1, x_2, x_3) is the predictor vector (i.e. tropical SSTAs) and \hat{y} is the SVM model output (i.e. HRV heavy rainfall frequency). The $\Phi(x_1, x_2, x_3)$ is the nonlinear kernel functions, and the w^T is the support vector denoting the norm of the nonlinear hyperpanel defined by $\Phi(x_1, x_2, x_3)$. In this study, we apply a least square SVM, meaning that the w^T and b in Eq. (8) are formulated to minimize the cost function (see details in Appendix 2).

Taking into account the nonlinear relationship between predictors and responses, the SVM regression substantially improves the prediction of heavy rainfall frequency over the HRV. The R^2 reaches 0.47, indicating that the SVM increases the explained variance in comparison with the multiple linear regression model (Fig. 7). More importantly, the regression slope better approaches the $y = x$ line than the linear model (Fig. 7). The results support our postulation and suggest that nonlinearity should be considered when constructing prediction models for HRV heavy rainfall using preceding tropical SSTA.

The probable sources of the nonlinearity might come from the nonlinear relationship between atmospheric moisture content and air temperature (Chou et al. 2009; Held and Soden 2006), as well as the potential positive feedback between heavy precipitation and vertical motion (Chou et al. 2012). These nonlinear processes are missing in linear models, but can be captured by the SVM algorithm built on nonlinear kernel functions. However, it is noteworthy that the explicit expression of SVM regression is virtually impossible to obtain, making it inapplicable to identify the mechanisms responsible for the nonlinearity. To achieve the

mechanistic understanding, numerical simulations should be implemented, which is beyond the scope of this study.

6 Conclusions

Summertime heavy rainfall events over the HRV in China are vitally important to the regional and national agriculture, economy and social development. However, accurately predicting HRV heavy rainfall remains challenging due to its complicated statistical behavior and the poorly understood physical processes governing heavy rainfall.

This study advances the understanding of and improves prediction skill of HRV heavy rainfall events by applying a novel rainfall framework built on a three-cluster *Normal* mixture model (Li and Li 2013). The three clusters reflect the probability behavior of light, moderate and heavy rainfall. Bayesian inference and the Gibbs sampler using Markov Chain Monte Carlo algorithm are applied to sample the distribution parameters of the model. Compared to traditionally used distribution models, the new framework improves the statistical inference on the HRV summer rainfall.

The analysis shows that heavy rainfall cluster contributes the largest amount of variance (58 %) to summer precipitation over the HRV, almost three times higher than that of light and moderate rainfall clusters. Furthermore, the contribution of heavy rainfall is manifested by the interannual variation of heavy rainfall frequency, whereas the variability of heavy rainfall intensity is secondary.

The HRV heavy rainfall frequency is most influenced by SSTA patterns over the north Indian Ocean, equatorial western Pacific, and the tropical Atlantic that collectively modulate summertime atmospheric circulation. When the tropical oceans are warmer than normal, abnormally stronger southerly wind supplies more moisture to the HRV, which favors the onset and maintenance of heavy rainfall events over the HRV.

The SSTA signals occur five months prior to the summer season, providing potential sources of prediction skill for heavy rainfall frequency over the HRV. Two statistical prediction models are thus constructed and tested: multiple linear regression model and the SVM algorithm. Both prediction models show considerable accuracy in predicting the frequency of HRV heavy rainfall events. Comparatively, the SVM algorithm further improves the predictions due to its capability to capture the nonlinear relationship between SSTA and rainfall over the HRV. Thus, our study suggests that the application of the new rainfall framework and the SVM algorithm has the potential to improve seasonal prediction of heavy rainfall frequency over the HRV region. The results obtained in this study have important implication to improve the nation's disaster early warning

system, which can help reduce economic and agriculture losses caused by heavy rainfall and related natural hazards.

Acknowledgments The authors thank Drs. David Dunson and David Bank for helpful discussion, and Dr. Georgy Manucharyan for comments on and editorial help with the manuscript. This work is supported by the NSF Grant AGS-1147608, NIH-1R21AG044294-01A1, and CAS XDA11010402. L. Li is partially supported by the WHOI postdoctoral scholarship, with funding provided by the Ocean and Climate Change Institute.

Appendix 1: Influence of data autocorrelation on Bayesian inference of *Normal* mixture model

Climate variables at daily scale usually contain certain autocorrelation, which violates the requirements by many statistical models. However, autocorrelation does not impact Bayesian inference of *Normal* mixture model, because it does not require statistical samples to be strictly independent and identically distributed (i.i.d.) (Hoff 2009).

Figure 8a shows the autocorrelation of HRV daily rainfall. According to the 52-year sample sets, the autocorrelation with 1-day lag is statistically significant ($\rho = 0.4$ for 52-year average, significant at 0.01 level), indicating that the rainfall events is highly dependent on those 1 day prior (Fig. 8a).

To eliminate the impact of data autocorrelation on Bayesian inference, we ran the framework (Eqs. 1–6) by using two subsets of de-autocorrelated data samples. In each subset, the rainfall data are sampled sequentially at 2-day interval, when autocorrelation becomes insignificant (Fig. 8a). The comparison between the inference on the de-autocorrelated sample sets and the original sample sets is shown in Fig. 8b, c. For the three rainfall clusters, neither cluster intensity nor cluster weight (frequency) is sensitive to the sample autocorrelation. The distribution parameters derived from the two sample sets are aligned tightly to the line of $y = x$, indicating the inference generates identical results using the two different sample sets (Fig. 8b, c).

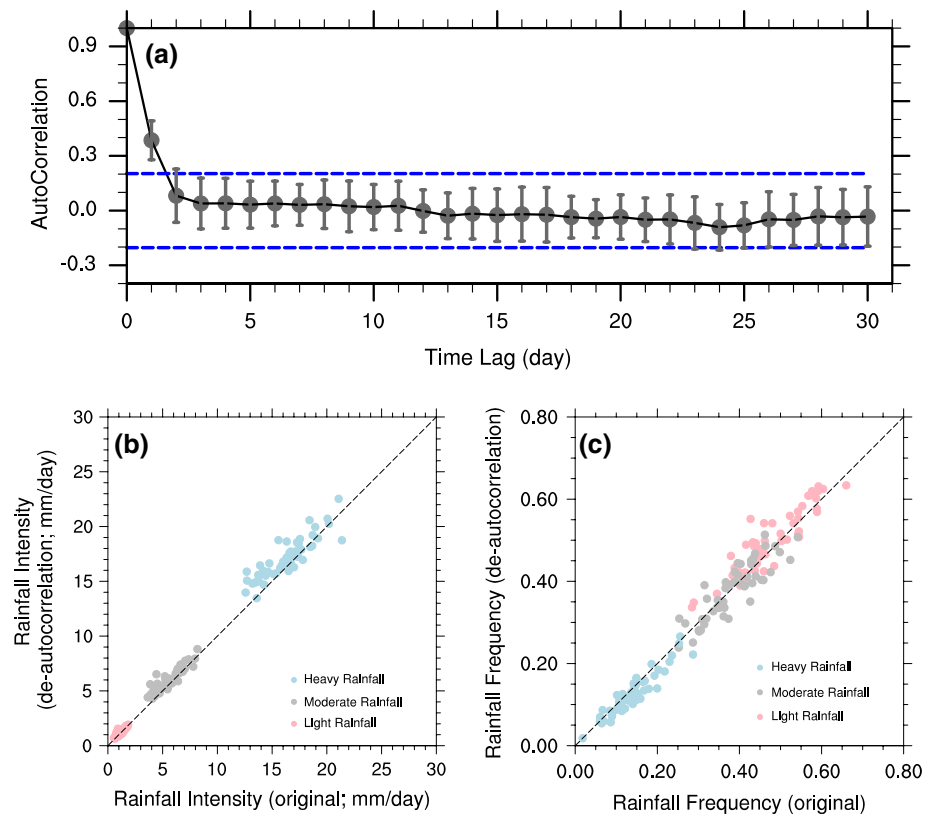
In conclusion, the Bayesian inference on *Normal* mixture model does not require the sample to be strictly i.i.d., making it easy to apply to climate variables, including precipitation and temperature, which usually have high autocorrelation.

Appendix 2: SVM algorithm

In this study, we apply a least square SVM, meaning that the w^T and b in Eq. (8) are formulated to minimize the cost function:

$$\psi_L(w, \varepsilon) = C \sum_{i=1}^n \varepsilon_i^2 + \frac{1}{2} \|w\|^2 \quad (9)$$

Fig. 8 **a** Autocorrelation of daily precipitation samples; the *dots* denote the maximum likelihood estimator of the autocorrelation coefficients, and the *error bars* denote one standard deviation of the autocorrelation derived from the daily precipitation during summer seasons in 1961–2012. The *dashed blue lines* are $\alpha = 0.05$ significance level. The autocorrelation located within the regions bounded by the two lines is statistically insignificant; **b**, **c** the comparison of Bayesian inference on HRV rainfall intensity and frequency using the original samples and those without autocorrelation



In the cost function, $\varepsilon_i^2 = (y_i - \hat{y}_i)^2$ is the quadratic loss term. C is a positive real constant reflecting the tolerance rate of prediction errors. The larger the value of C , the lower tolerance for prediction errors, and thus higher punishment added in the cost function. The primal SVM problem becomes an optimization problem to obtain $\min_{w \in R^n, b \in R} C \sum_{i=1}^n \varepsilon_i^2 + \frac{1}{2} \|w\|^2$. The solutions to the optimization problem can be obtained using the Lagrangian:

$$L(w, b, \vec{\varepsilon}, \vec{\alpha}) = \frac{1}{2} w^T w + C \sum_{i=1}^n \varepsilon_i^2 - \sum_{i=1}^n \alpha_i \{\hat{y}_i + \varepsilon_i - y_i\}^2,$$

where α_i are Lagrangian multipliers. The optimal solutions must satisfy the conditions [also known as Karush–Kuhn–Tucker (KKT) conditions]:

$$\begin{cases} \frac{\partial L}{\partial w} = w - \sum_{i=1}^n \alpha_i \Phi(\vec{x}_i) = 0 \\ \frac{\partial L}{\partial b} = \sum_{i=1}^n \alpha_i = 0 \\ \frac{\partial L}{\partial \varepsilon_i} = \alpha_i - C \varepsilon_i = 0 \\ \frac{\partial L}{\partial \alpha_i} = \hat{y}_i + \varepsilon_i - y_i = 0 \end{cases} \quad (10)$$

By solving the above set of linear equations, the SVM regression model can be obtained:

$$f(\vec{x}) = \sum_{ij} \alpha^* K(x_i, x_j) + b^*, \quad (11)$$

where α^* and b^* are the solutions to Eq. (10); and $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$.

References

- Bao M (2008) Relationship between persistent heavy rain events in the Huaihe River Valley and the distribution pattern of convective activities in the tropical western Pacific warm pool. *Adv Atmos Sci* 25:329–338
- Burges CJC (1998) A tutorial on support vector machines for pattern recognition. *Data Min Knowl Discov* 2:121–167
- Chang C-P, Zhang Y, Li T (2000) Interannual and interdecadal variations of the East Asian summer monsoon and tropical Pacific SSTs. Part I: roles of the subtropical ridge. *J Clim* 13:4310–4325
- Chen H, Sun J, Fan K (2012) Decadal features of heavy rainfall events in eastern China. *Acta Meteorol Sin* 26:289–303
- Cho H-K, Bowman KP, North GR (2004) A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission. *J Appl Meteorol* 43:1586–1597
- Chou C, Neelin JD, Chen C-A, Tu J-Y (2009) Evaluating the “rich-get-richer” mechanism in tropical precipitation change under global warming. *J Clim* 22:1982–2005
- Chou C, Chen C-A, Tan P-H, Chen KT (2012) Mechanisms for global warming impacts on precipitation frequency and intensity. *J Clim* 25:3291–3306
- Christensen R (2011) Plain answers to complex questions: the theory of linear models, vol 465. Springer, New York
- Ding Y (1992) Summer monsoon rainfalls in China. *J Meteorol Soc Japan* 70:373–396

- Gelfand AE (2000) Gibbs sampling. *J Am Stat Assoc* 95:1300–1304
- Gill AE (1980) Some simple solutions for the heat induced tropical circulation. *Q J R Meteorol Soc* 106:447–462
- Gumble EJ (ed) (1954) Statistical theory of extreme values and some practical applications, vol 33. Department of commerce, National Bureau of Standards, Washington
- Held IM, Soden BJ (2006) Robust responses of the hydrological cycle to global warming. *J Clim* 19:5686–5699
- Higgins RW, Silva VBS, Shi W, Larson J (2007) Relationships between climate variability and fluctuations in daily precipitation over the United States. *J Clim* 20:3561–3579
- Hoff PD (2009) A first course in Bayesian statistical methods. Springer, Berlin
- Hong J, Liu Y (2012) Contrasts of atmospheric circulation and associated tropical convection between Huaihe River Valley and Yangtze River Valley Mei-yu flooding. *Adv Atmos Sci* 29:755–768
- Huang B, Shukla J (2005) Ocean–atmosphere interactions in the tropical and subtropical Atlantic Ocean. *J Clim* 18:1652–1672
- Huang R, Sun F (1992) Impacts of the tropical Western Pacific on the East Asian summer monsoon. *J Meteorol Soc Japan* 70(1B):243–256
- Ji X, Neelin JD, Lee S-K, Mechoso CR (2014) Interhemispheric teleconnections from tropical heat sources in intermediate and simple models. *J Clim* 27:684–697
- Kalnay E et al (1996) The NCEP–NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77:437–471
- Katz RW, Brown BG (1992) Extreme events in a changing climate: variability is more important than averages. *Clim Change* 21:289–302
- Kosaka Y, Nakamura H (2010) Mechanisms of meridional teleconnection observed between a summer monsoon system and a subtropical anticyclone. Part II: a global survey. *J Clim* 23:5109–5125
- Kosaka Y, Xie S-P, Lau N-C, Vecchi GA (2013) Origin of seasonal predictability for summer climate over the Northwestern Pacific. *Proc Natl Acad Sci* 110:7574–7579
- Lau K-M (1992) East Asian summer monsoon rainfall variability and climate teleconnection. *J Meteorol Soc Japan* 70:211–241
- Lei Y, Hoskins B, Slingo J (2011) Exploring the interplay between natural decadal variability and anthropogenic climate change in summer rainfall over China. Part I: observational evidence. *J Clim* 24:4584–4599
- Li Y, Leung LR (2013) Potential impacts of the Arctic on interannual and interdecadal summer precipitation over China. *J Clim* 26:899–917
- Li L, Li W (2013) Southeastern United States summer rainfall framework and its implication for seasonal prediction. *Environ Res Lett* 8:044017
- Li Y et al (2011) Effects of climate variability and change on Chinese agriculture: a review. *Clim Res* 50:83–102
- Li W, Li L, Ting M, Liu Y (2012) Intensification of Northern Hemisphere subtropical highs in a warming climate. *Nat Geosci* 5:830–834
- Lin TI, Lee JC, Yan SY (2007) Finite mixture modeling using the skew normal distribution. *Stat Sin* 17:909–927
- Liu Y, Wu G, Ren R (2004) Relationship between the subtropical anticyclone and diabatic heating. *J Clim* 17:682–698
- Liu Y, Hoskins BJ, Blackburn M (2007) Impacts of the Tibetan topography and heating on the summer flow over Asia. *J Meteorol Soc Japan* 85B:1–19
- Matsumoto J (1988) Heavy rainfall over East Asia. *Int J Climatol* 9:407–423
- McLachlan GP, Peel D (2000) Finite mixture models. Wiley, New York
- Melnykov V, Maitra R (2010) Finite mixture models and model-based clustering. *Stat Surv* 4:80–116
- Meng L, Long D, Quiring SM, Shen Y (2014) Statistical analysis of the relationship between spring soil moisture and summer precipitation in East China. *Int J Climatol* 34:1511–1523
- Qian W, Fu J, Yan Z (2007) Decrease of light rain events in summer associated with a warming environment in China during 1961–2005. *Geophys Res Lett* 34:L11705
- Richardson S, Green PJ (1997) On Bayesian analysis of mixtures with an unknown number of components. *J R Stat Soc B* 59:731–792
- Ropelewski CF, Halpert MS (1987) Global and regional scale precipitation patterns associated with the El Niño/Southern oscillation. *Mon Weather Rev* 115:1606–1626
- Smith TM, Reynolds RW, Peterson TC, Lawrimore J (2008) Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006). *J Clim* 21:2283–2296
- Stephens M (2000) Dealing with label switching in mixture models. *J R Stat Soc B* 62:795–809
- Sun J, Wang H (2012) Changes of the connection between the summer North Atlantic Oscillation and the East Asian summer rainfall. *J Geophys Res Atmos* 117:D08110
- Ting M, Wang H (1997) Summertime U.S. precipitation variability and its relation to Pacific sea surface temperature. *J Clim* 10:1853–1873
- Vihma T (2014) Effects of Arctic Sea Ice decline on weather and climate: a review. *Surv Geophys*. doi:10.1007/s10712-0014-19284-10710
- Wang B, Wu R, Li T (2003) Atmosphere–warm ocean interaction and its impacts on Asian–Australian monsoon variation. *J Clim* 16:1195–1211
- Wang T, Wang H, Otterå OH, Gao YQ, Suo LL, Furevik T, Yu L (2013) Anthropogenic agent implicated as a prime driver of shift in precipitation in eastern China in the late 1970s. *Atmos Chem Phys* 13:12433–12450
- Wilson PS, Toumi R (2005) A fundamental probability distribution for heavy rainfall. *Geophys Res Lett* 32:L14812
- Wu F, Fu C (2013) Change of precipitation intensity spectra at different spatial scales under warming conditions. *Chin Sci Bull* 58:1385–1394
- Wu R, Wang B (2002) A contrast of the East Asian summer monsoon–ENSO relationship between 1962–77 and 1978–93. *J Clim* 15:3266–3279
- Xie P, Chen M, Yang S, Yatagai A, Hayasaka T, Fukushima Y, Liu C (2007) A gauge-based analysis of daily precipitation over East Asia. *J Hydrometeorol* 8:607–626
- Yang F, Lau K-M (2004) Trend and variability of China precipitation in spring and summer: linkage to sea-surface temperatures. *Int J Climatol* 24:1625–1644
- Yang J, Liu Q, Liu Z (2010) Linking observations of the Asian monsoon to the Indian Ocean SST: possible roles of Indian Ocean basin mode and dipole mode. *J Clim* 23:5889–5902
- Zhang R (2001) Relations of water vapor transport from Indian monsoon with that over East Asia and the summer rainfall in China. *Adv Atmos Sci* 18:1006–1017
- Zhang Y-L, You W-J (2014) Social vulnerability to floods: a case study of Huaihe River Basin. *Nat Hazards* 71:2113–2125
- Zhang R, Zuo Z (2011) Impact of spring soil moisture on surface energy balance and summer monsoon circulation over East Asia and precipitation in East China. *J Clim* 24:3309–3322
- Zhou T, Yu R (2005) Atmospheric water vapor transport associated with typical anomalous summer rainfall patterns in China. *J Geophys Res Atmos* 110:D08104
- Zhou T et al (2009) Why the western Pacific subtropical high has extended westward since the late 1970s. *J Clim* 22:2199–2215
- Zuo J, Li W, Sun C, Xu L, Ren H-L (2013) Impact of the North Atlantic sea surface temperature tripole on the East Asian summer monsoon. *Adv Atmos Sci* 30:1173–1186